

Phi-Divergence test statistics for testing the validity of latent class models for binary data

Felipe¹, A., Martin², N., Miranda¹, P. and Pardo¹, L.

¹Department of Statistics and O.R., Complutense University of Madrid, 28040-Madrid, Spain

²Department of Statistics, Carlos III University of Madrid, 28903-Getafe (Madrid), Spain

Abstract

The main purpose of this paper is to present new families of test statistics for studying the problem of goodness-of-fit of some data to a latent class model for binary data. The families of test statistics introduced are based on phi-divergence measures, a natural extension of maximum likelihood. We also treat the problem of testing a nested sequence of latent class models for binary data. For these statistics, we obtain their asymptotic distribution. Finally, a simulation study is carried out in order to compare the efficiency, in the sense of the level and the power, of the new statistics considered in this paper for sample sizes that are not big enough to apply the asymptotical results.

MSC: Primary 62F03; 62F05; Secondary 62H15

Keywords: Latent class models, Minimum phi-divergence estimator, Maximum likelihood estimator, Asymptotic distribution, Phi-divergence test statistics, Nested latent class models

1 Introduction

Consider a set \mathcal{P} of N people: $\mathcal{P} := \{P_1, \dots, P_N\}$. Each person P_v is asked to answers to k dichotomous items I_1, \dots, I_k ; let us denote by y_{vi} the answer of person P_v to item I_i , i.e.

$$y_{vi} := \begin{cases} 1 & \text{if the answer of } P_v \text{ to } I_i \text{ is correct} \\ 0 & \text{otherwise} \end{cases}.$$

Let $\mathbf{y}_v := (y_{v1}, \dots, y_{vk})$ denote a generic pattern of right and wrong answers to the k items given by person P_v . In order to explain the statistical relationships among the observed variables, a categorical latent variable (categorical unobservable variable) is postulated to exist, whose different levels partition set \mathcal{P} into m mutually exclusive and exhaustive *latent classes*. Let us denote these classes by C_1, \dots, C_m and their corresponding relative sizes by w_1, \dots, w_m ; thus, w_j denotes the probability of a randomly selected person $P_v \in \mathcal{P}$ belongs to class C_j , i.e.

$$w_j = \Pr(P_v \in C_j), j = 1, \dots, m.$$

We denote by p_{ji} the probability of a right answer of P_v to the item I_i under the assumption that P_v is in class C_j :

$$p_{ji} = \Pr(y_{vi} = 1 | P_v \in C_j), j = 1, \dots, m, i = 1, \dots, k.$$

Let \mathbf{y}_ν be a possible answer vector. We shall assume that in each class the answers for the different questions are stochastically independent; therefore, we can write

$$Pr(\mathbf{y}_\nu | P_v \in C_j) = \prod_{i=1}^k p_{ji}^{y_{\nu i}} (1 - p_{ji})^{1-y_{\nu i}},$$

and

$$Pr(\mathbf{y}_\nu) = \sum_{j=1}^m w_j \prod_{i=1}^k p_{ji}^{y_{\nu i}} (1 - p_{ji})^{1-y_{\nu i}}. \quad (1)$$

There are 2^k possible answer vectors \mathbf{y}_ν whose probability of occurrence are given by Eq. (1); they constitute the *manifest probabilities* for the items I_1, \dots, I_k in the population given by P_1, \dots, P_N . The probability vector $(Pr(\mathbf{y}_1), \dots, Pr(\mathbf{y}_{2^k}))$ characterizes a latent class model (LCM) for binary data.

We will denote by N_ν , $\nu = 1, \dots, 2^k$, the number of times that the sequence \mathbf{y}_ν appears in an N -sample and

$$\hat{\mathbf{p}} := (N_1/N, \dots, N_{2^k}/N).$$

The likelihood function L is given by

$$L(w_1, \dots, w_m, p_{11}, \dots, p_{mk}) = Pr(N_1 = n_1, \dots, N_{2^k} = n_{2^k}) = \frac{N!}{\prod_{\nu=1}^{2^k} n_\nu!} \prod_{\nu=1}^{2^k} Pr(\mathbf{y}_\nu)^{n_\nu}. \quad (2)$$

By n_ν we are denoting a realization of the random variable N_ν , $\nu = 1, \dots, 2^k$. In this model the unknown parameters are w_j , $j = 1, \dots, m$ and p_{ji} , $j = 1, \dots, m$, $i = 1, \dots, k$. These parameters can be estimated using the maximum likelihood estimator (e.g. McHugh (1956), Lazarsfeld & Henry (1968), Clogg (1995)). In order to avoid the problem of obtaining uninterpretable estimations for the item latent probabilities lying outside the interval $[0, 1]$, some authors (Lazarsfeld & Henry (1968), Formann (1976), Formann (1977), Formann (1978), Formann (1982), Formann (1985)) proposed a linear-logistic parametrization for w_j and p_{ji} given by

$$p_{ji} = \frac{\exp(x_{ji})}{1 + \exp(x_{ji})}, \quad j = 1, \dots, m, \quad i = 1, \dots, k,$$

and

$$w_j = \frac{\exp(z_j)}{\sum_{h=1}^m \exp(z_h)}, \quad j = 1, \dots, m.$$

Next, restrictions are introduced relating parameters x_{ji} , w_j to some explanatory variables, defined through parameters λ_r , $r = 1, \dots, t$ and η_s , $s = 1, \dots, u$, so the final model is given by

$$p_{ji} = \frac{\exp(\sum_{r=1}^t q_{jir} \lambda_r + c_{ji})}{1 + \exp(\sum_{r=1}^t q_{jir} \lambda_r + c_{ji})}, \quad j = 1, \dots, m, \quad i = 1, \dots, k, \quad (3)$$

and

$$w_j = \frac{\exp(\sum_{r=1}^u v_{jr} \eta_r + d_j)}{\sum_{h=1}^m \exp(\sum_{r=1}^u v_{hr} \eta_r + d_h)}, \quad j = 1, \dots, m, \quad (4)$$

where

$$\mathbf{Q}_r = (q_{jir})_{\substack{j=1, \dots, m \\ i=1, \dots, k}}, \quad r = 1, \dots, t, \quad \mathbf{C} = (c_{ji})_{\substack{j=1, \dots, m \\ i=1, \dots, k}}, \quad \mathbf{V} = (v_{jr})_{\substack{j=1, \dots, m \\ r=1, \dots, u}}, \quad \mathbf{d} = (d_j)_{j=1, \dots, m},$$

are fixed. Matrix \mathbf{Q} specifies to which amount the predictors defined through parameters λ_r are relevant for each x_{ji} . The terms c_{ji} were introduced to include the possibility that certain p_{ji} are fixed to certain previously determined values; this possibility was considered by Goodman (1974). The same applies for matrix \mathbf{V} : thus, \mathbf{V} specifies to which amount η_s is relevant for each z_j . The terms d_j are introduced to include the possibility that certain z_j are fixed to certain previously determined values.

Consequently, in this case the vector of unknown parameters $\boldsymbol{\theta}$ in the LCM for binary data is given by

$$\boldsymbol{\theta} := (\boldsymbol{\lambda}, \boldsymbol{\eta}),$$

where $\boldsymbol{\lambda}$ and $\boldsymbol{\eta}$ are defined as

$$\boldsymbol{\lambda} := (\lambda_1, \dots, \lambda_t), \quad \boldsymbol{\eta} := (\eta_1, \dots, \eta_u).$$

Once $\boldsymbol{\lambda}$ and $\boldsymbol{\eta}$ are estimated, relations (3) and (4) give estimations for the parameters w_j ($j = 1, \dots, m$) and p_{ji} ($j = 1, \dots, m; i = 1, \dots, k$)

By $\boldsymbol{\Theta}$ we shall denote the set in which the parameter $\boldsymbol{\theta}$ varies, i.e. the parametric space. Thus, we have $t + u$ unknown parameters that can be estimated by maximum likelihood through Eq. (2).

In Felipe et al. (2014), a new procedure for estimating p_{ji} , w_j , $i = 1, \dots, k$, $j = 1, \dots, m$, estimating previously the parameters λ_i , $i = 1, \dots, t$ and η_j , $j = 1, \dots, u$ was presented. It consists in introducing in the context of LCM for binary data a new family of estimators based on divergence measures: *Minimum ϕ -divergence estimators* (M ϕ E). As shown in Felipe et al. (2014), this family of estimators contains as a particular case the classical maximum likelihood estimator (MLE). M ϕ E were introduced for the first time in Morales et al. (1995) and since then, many interesting estimation problems have been solved using them (see e.g. Pardo (2006)).

Let us briefly explain this procedure. Consider two probability distributions $\mathbf{p} = (p_1, \dots, p_M)$, $\mathbf{q} = (q_1, \dots, q_M)$ and a function ϕ that is convex for $x > 0$ and satisfies $\phi(1) = 0$, $0\phi(0/0) = 0$ and

$$0\phi(p/0) = p \lim_{x \rightarrow \infty} \frac{\phi(x)}{x}.$$

The ϕ -divergence measure between the probability distributions \mathbf{p} and \mathbf{q} is defined by

$$D_\phi(\mathbf{p}, \mathbf{q}) := \sum_{i=1}^M p_i \phi\left(\frac{q_i}{p_i}\right).$$

Given a LCM for binary data with parameters $\boldsymbol{\lambda} = (\lambda_1, \dots, \lambda_t)$ and $\boldsymbol{\eta} = (\eta_1, \dots, \eta_u)$, the M ϕ E of $\boldsymbol{\theta} = (\boldsymbol{\lambda}, \boldsymbol{\eta})$ is any $\hat{\boldsymbol{\theta}}_\phi$ satisfying

$$\hat{\boldsymbol{\theta}}_\phi = \arg \min_{(\boldsymbol{\lambda}, \boldsymbol{\eta}) \in \boldsymbol{\Theta}} D_\phi(\hat{\mathbf{p}}, \mathbf{p}(\boldsymbol{\lambda}, \boldsymbol{\eta})) \quad (5)$$

where $D_\phi(\hat{\mathbf{p}}, \mathbf{p}(\boldsymbol{\lambda}, \boldsymbol{\eta}))$ is the ϕ -divergence measure between the probability vectors $\hat{\mathbf{p}}$ and $\mathbf{p}(\boldsymbol{\lambda}, \boldsymbol{\eta})$, given by

$$D_\phi(\hat{\mathbf{p}}, \mathbf{p}(\boldsymbol{\lambda}, \boldsymbol{\eta})) = \sum_{\nu=1}^{2^k} p(\mathbf{y}_\nu, \boldsymbol{\lambda}, \boldsymbol{\eta}) \phi\left(\frac{\hat{p}_\nu}{p(\mathbf{y}_\nu, \boldsymbol{\lambda}, \boldsymbol{\eta})}\right). \quad (6)$$

For more details about ϕ -divergence measures see Cressie and Pardo (2002) and Pardo (2006). In the particular case of $\phi(x) = x \log x - x + 1$, we obtain the so-called *Kullback-Leibler divergence measure*, i.e.

$$D_{Kullback}(\hat{\mathbf{p}}, \mathbf{p}(\boldsymbol{\lambda}, \boldsymbol{\eta})) = \sum_{\nu=1}^{2^k} \hat{p}_\nu \log \frac{\hat{p}_\nu}{p(\mathbf{y}_\nu, \boldsymbol{\lambda}, \boldsymbol{\eta})}. \quad (7)$$

It is not difficult to establish (see Felipe et al (2014)) that

$$\log L(w_1, \dots, w_m, p_{11}, \dots, p_{mk}) = -ND_{Kullback}(\hat{\mathbf{p}}, \mathbf{p}(\boldsymbol{\lambda}, \boldsymbol{\eta})) + \text{constant},$$

Therefore, maximizing Eq. (2) in λ and η is equivalent to minimizing Eq. (7) in λ and η . Consequently, the value $\hat{\theta} = (\hat{\lambda}, \hat{\eta})$ that minimizes $\theta = (\lambda, \eta)$ in the Kullback-Leibler divergence is the MLE of the parameters for the LCM for binary data or equivalently, the minimum Kullback-Leibler divergence estimator. We shall denote it by $\hat{\theta}$ or

$$\hat{\theta}_{Kullback} := \arg \min_{(\lambda, \eta) \in \Theta} D_{Kullback}(\hat{p}, p(\lambda, \eta)).$$

This fact allows us to say that $M\phi E$ is a natural extension of the MLE.

The rest of the paper is organized as follows: In Section 2 we study the problem of goodness-of-fit when dealing with phi-divergence measures; two families of test statistics generalizing the classical ones studied in Formann (1985) are introduced and their asymptotical behavior is established. In Section 3, we proceed the same way for the problem of determining the best model in a nested sequence; besides, we also provide the asymptotical behavior of the test statistics. Section 4 is devoted to a simulation study. We finish with the conclusions. In an appendix we give the proofs of the results presented in Sections 2 and 3.

2 Goodness-of-fit tests

LCM for binary data fit is assessed by comparing the observed classification frequencies to the expected frequencies predicted by the LCM for binary data. When dealing with the MLE, the difference is formally assessed with a likelihood ratio test statistic or with a chi-square test statistic whose expressions are given by

$$G^2 = 2N \sum_{\nu=1}^{2^k} \hat{p}_{\nu} \log \frac{\hat{p}_{\nu}}{p(\mathbf{y}_{\nu}, \hat{\lambda}, \hat{\eta})} \quad (8)$$

and

$$X^2 = \sum_{\nu=1}^{2^k} \frac{(n_{\nu} - Np(\mathbf{y}_{\nu}, \hat{\lambda}, \hat{\eta}))^2}{Np(\mathbf{y}_{\nu}, \hat{\lambda}, \hat{\eta})}, \quad (9)$$

respectively.

It is well-known that the asymptotic distribution of the test statistics X^2 and G^2 is a chi-square distribution with $2^k - (u + t) - 1$ degrees of freedom, see Forman (1985). It is a simple exercise to see that these test statistics are particular cases of the more general family of test statistics

$$T_{\phi} = \frac{2N}{\phi''(1)} D_{\phi}(\hat{p}, p(\hat{\lambda}, \hat{\eta})) \quad (10)$$

taking $\phi(x) = \frac{1}{2}(x-1)^2$ and $\phi(x) = x \log x - x + 1$, respectively. In the following we shall denote $\hat{\theta} = (\hat{\lambda}, \hat{\eta})$ and we shall write $D_{\phi}(\hat{p}, p(\hat{\theta}))$. Therefore, Eq. (10) gives a family of test statistics for the problem of goodness-of-fit to some data to a LCM. In Eq. (10) parameters λ and η are estimated using the MLE, but notice that MLE is a particular case of the $M\phi E$.

Based on the $M\phi E$ defined in Eq. (5), we shall consider in this paper the phi-divergence family of test statistics given by

$$T_{\phi_1}^{\phi_2} := \frac{2N}{\phi_1''(1)} D_{\phi_1}(\hat{p}, p(\hat{\theta}_{\phi_2})) \quad (11)$$

where $\hat{\theta}_{\phi_2} = (\hat{\lambda}_{\phi_2}, \hat{\eta}_{\phi_2})$.

This family of test statistics is a natural extension of the family (10) in which the MLE has been replaced by the $M\phi_2 E$. Notice that in the family presented in (11) we have the possibility to use one measure of divergence based on a function ϕ_2 for the problem estimation and another measure of divergence based on a function ϕ_1 for the problem of testing.

In the following theorem we present the asymptotic distribution of this family of test statistics.

Theorem 1 Under the hypothesis that the LCM for binary data with parameters $\lambda = (\lambda_1, \dots, \lambda_t)$ and $\eta = (\eta_1, \dots, \eta_u)$ holds, the asymptotic distribution of the family of test statistics $T_{\phi_1}^{\phi_2}$ given in (11) is a chi-square distribution with $2^k - (u + t) - 1$ degrees of freedom.

Proof. See Appendix ■

It is noteworthy that the asymptotical distribution does not depend on ϕ , i.e. it is the same for any function ϕ considered.

Let us see an example:

Example 2 We consider the interview data collected by Coleman (1964) and analyzed later in Goodman (1974); this model is explained in Formann (1982) and Formann (1985). The experiment consists in evaluating the answers of 3398 schoolboys to two questions about their membership in the “leading crowd” on two occasions t_1 and t_2 (October, 1957 and May, 1958). Thus, in this model we have 4 questions and there are four manifest variables (answers to both questions at both moments); these answers can only be “low” (value 0) and “high” (value 1), so that the manifest variables are dichotomous. The sample data is given in next table:

October, 1957/ May, 1958	00	01	10	11
00	554	338	97	85
01	281	531	75	184
10	87	56	182	171
11	49	110	140	458

Next, 4 latent classes are considered, namely

$C_1 \equiv$ low agreement in question 1 and low agreement in question 2.

$C_2 \equiv$ low agreement in question 1 and high agreement in question 2.

$C_3 \equiv$ high agreement in question 1 and low agreement in question 2.

$C_4 \equiv$ high agreement in question 1 and high agreement in question 2.

There are 16 probability values p_{ji} to be estimated; we consider the first hypothesis appearing in Formann (1985), namely “The attitudinal changes between times t_1 and t_2 are dependent on the positions (low, high) of the respective classes on the underlying attitudinal scales at t_1 ”. Thus, a model with 8 parameters λ_i is considered; λ_1 means low agreement in the first question at time t_1 , λ_2 means high agreement in the first question at time t_1 , λ_3 means low agreement in the second question at time t_1 , λ_4 means high agreement in the second question at time t_1 , and $\lambda_5, \lambda_6, \lambda_7, \lambda_8$ are the same parameters at time t_2 . We write the values for matrices \mathbf{Q}_i as they appear in Formann (1985). In our notation, the matrices \mathbf{Q}_i can be derived considering the i -th column in the table and dividing it in four columns of four elements each (each corresponding to a latent class).

Class	Item	λ_1	λ_2	λ_3	λ_4	λ_5	λ_6	λ_7	λ_8
1	1	1	0	0	0	0	0	0	0
	2	0	0	1	0	0	0	0	0
	3	0	0	0	0	1	0	0	0
	4	0	0	0	0	0	0	1	0
2	1	1	0	0	0	0	0	0	0
	2	0	0	0	1	0	0	0	0
	3	0	0	0	0	1	0	0	0
	4	0	0	0	0	0	0	0	1
3	1	0	1	0	0	0	0	0	0
	2	0	0	1	0	0	0	0	0
	3	0	0	0	0	0	1	0	0
	4	0	0	0	0	0	0	1	0
4	1	0	1	0	0	0	0	0	0
	2	0	0	0	1	0	0	0	0
	3	0	0	0	0	0	1	0	0
	4	0	0	0	0	0	0	0	1

Note that the hypothesis is that the attitudinal changes between times t_1 and t_2 are dependent upon the items as well as on the classes. For this reason, the part corresponding each latent class can be partitioned in four submatrices of size 2×4 . The submatrices lying on the main diagonal are the same by the hypothesis defining the model and the two other submatrices are null. The differences among them are due to the differences in the latent classes. Next, $c_{ij} = 0, \forall i, j$ (as we have explained when values c_{ij} were introduced in Section 1). Finally, 4 parameters η_j are considered, taking as matrix \mathbf{V} the identity matrix and $d_j = 0, \forall j$.

It is noteworthy that our model assumes that answers to the questions are conditionally independent given the latent class. In this example, we are dealing with repeated responses to two questions, so this assumption may be unrealistic. However, this assumption is made in the original paper of Goodman (1974) and we follow this assumption for the sake of the example.

In order to study if the data are from a LCM for binary data we shall consider the particular family of phi-divergence measures introduced and studied by Cressie and Read (1984): The power divergence family. This family is obtained from

$$\phi(x) \equiv \phi_a(x) = \begin{cases} \frac{1}{a(a+1)}(x^{a+1} - x - a(x-1)) & a \neq 0, a \neq -1 \\ x \log x - x + 1 & a = 0 \\ -\log x + x - 1 & a = -1 \end{cases} \quad (12)$$

In Felipe et al. (2014) it was established, on the basis of a simulation study, that a good alternative to the MLE is the M ϕ E obtained from Eq. (5) with $a = 2/3$, i.e.,

$$\hat{\boldsymbol{\theta}}_{2/3} = \arg \min_{(\boldsymbol{\lambda}, \boldsymbol{\eta}) \in \Theta} D_{2/3}(\hat{\mathbf{p}}, \mathbf{p}(\boldsymbol{\lambda}, \boldsymbol{\eta})),$$

being

$$D_{2/3}(\hat{\mathbf{p}}, \mathbf{p}(\boldsymbol{\lambda}, \boldsymbol{\eta})) = \frac{9}{10} \left(\sum_{j=1}^{2^k} \frac{\hat{p}_j^{5/3}}{p_j(\boldsymbol{\lambda}, \boldsymbol{\eta})^{2/3}} - 1 \right)$$

Therefore we are going to consider in our study the M ϕ E obtained with $\phi(x)$ defined in Eq. (12) for $a = 2/3$ in order to get an estimation of parameters $\boldsymbol{\lambda}$ and $\boldsymbol{\eta}$. In Table 1 we present the values obtained for these parameters, as well as the estimation of the probabilities and the weights of the latent classes

Parameter / a		Parameter / a	
$\hat{\lambda}_1$	-2.34292610	$\hat{p}_{1,1}$	0.08762969
$\hat{\lambda}_2$	1.72393168	$\hat{p}_{1,2}$	0.30144933
$\hat{\lambda}_3$	-0.84040580	$\hat{p}_{1,3}$	0.11256540
$\hat{\lambda}_4$	1.56524945	$\hat{p}_{1,4}$	0.28671773
$\hat{\lambda}_5$	-2.06480043	$\hat{p}_{2,1}$	0.08762969
$\hat{\lambda}_6$	2.29928080	$\hat{p}_{2,2}$	0.82710532
$\hat{\lambda}_7$	-0.91137901	$\hat{p}_{2,3}$	0.11256540
$\hat{\lambda}_8$	2.01252338	$\hat{p}_{2,4}$	0.88210569
$\hat{\eta}_1$	0.50480183	$\hat{p}_{3,1}$	0.8463457
$\hat{\eta}_2$	0.16964329	$\hat{p}_{3,2}$	0.30144933
$\hat{\eta}_3$	-0.87356633	$\hat{p}_{3,3}$	0.90881746
$\hat{\eta}_4$	-0.00424661	$\hat{p}_{3,4}$	0.28671773
\hat{w}_1	0.38936544	$\hat{p}_{4,1}$	0.84863457
\hat{w}_2	0.27848377	$\hat{p}_{4,2}$	0.82710532
\hat{w}_3	0.09811597	$\hat{p}_{4,3}$	0.90881746
\hat{w}_4	0.23403482	$\hat{p}_{4,4}$	0.88210569

Table 1: Estimations of the parameters for Example 4.

Now we are interested in studying the goodness-of-fit of our data to this model. We shall consider the family of test statistics, $T_{\phi_a}^{\phi_{2/3}}$, obtained from $\phi_a(x)$ with $a = -1, -1/2, 0, 2/3, 1, 1.5, 2, 2.5$ and 3, i.e.,

$$T_{\phi_a}^{\phi_{2/3}} = \begin{cases} \frac{2N}{a(a+1)} \left(\sum_{\nu=1}^{2^k} \frac{\hat{p}_\nu^{a+1}}{p(\mathbf{y}_\nu, \hat{\boldsymbol{\theta}}_{2/3})^a} - 1 \right) & \text{if } a \neq 0, -1 \\ 2 \sum_{\nu=1}^{2^k} n_\nu \log \frac{\hat{p}_\nu}{p(\mathbf{y}_\nu, \hat{\boldsymbol{\theta}}_{2/3})} & a = 0 \\ 2N \sum_{\nu=1}^{2^k} p(\mathbf{y}_\nu, \hat{\boldsymbol{\theta}}_{2/3}) \log \frac{p(\mathbf{y}_\nu, \hat{\boldsymbol{\theta}}_{2/3})}{\hat{p}_\nu} & a = -1 \end{cases} \quad (13)$$

The results are presented in the following table

a	-1	-1/2	0	2/3	1	3/2	2	5/2	3
$T_{\phi_a}^{\phi_{2/3}}$	1.279	1.278	1.277	1.277	1.277	1.277	1.278	1.279	1.281

On the other hand, the distribution of this statistics is a χ^2 with $16-11-1=4$ degrees of freedom; as $\chi_{4;0.05}^2 = 9.49$, we conclude that we have no evidence to reject our model.

Notice that the values for all test statistics are very similar; this was expected, as the sample size under consideration is big enough ($N = 3398$) to apply the asymptotical result of Theorem 1.

Remark 3 There are some classical measures of divergence which cannot be expressed as a ϕ -divergence measure, such as the divergence measures of Bhattacharya (1943), Rényi (1961), and Sharma and Mittal (1977). However, such measures are particular cases of the (h, ϕ) -divergence measures and can be defined by

$$D_{\phi_1}^h(\hat{\mathbf{p}}, \mathbf{p}(\hat{\boldsymbol{\lambda}}_{\phi_2}, \hat{\boldsymbol{\eta}}_{\phi_2})) := h\left(D_{\phi_1}(\hat{\mathbf{p}}, \mathbf{p}(\hat{\boldsymbol{\lambda}}_{\phi_2}, \hat{\boldsymbol{\eta}}_{\phi_2}))\right),$$

where h is a differentiable increasing function mapping from $[0, \infty)$ onto $[0, \infty)$, with $h(0) = 0$ and $h'(0) > 0$. In Table 2, these divergence measures are presented, along with the corresponding expressions of h and ϕ .

Divergence	$h(x)$	$\phi(x)$
Rényi	$\frac{1}{a(a-1)} \log(a(a-1)x + 1), \quad a \neq 0, 1$	$\frac{x^a - a(x-1) - 1}{a(a-1)}, \quad a \neq 0, 1$
Sharma-Mittal	$\frac{1}{b-1} \left\{ [1 + a(a-1)x]^{\frac{b-1}{a-1}} - 1 \right\}, \quad b, a \neq 1$	$\frac{x^a - a(x-1) - 1}{a(a-1)}, \quad a \neq 0, 1$
Battacharya	$-\log(-x + 1)$	$-x^{1/2} + \frac{1}{2}(x + 1)$

Table 2: Some specific (h, ϕ) -divergence measures.

The (h, ϕ) -divergence measures were introduced in Menéndez et al. (1995) and some associated asymptotic results for them were established in Menéndez et al. (1997). Moreover, some interesting results about Rényi divergence measures can be seen in Gil et al. (2013), Golshani et al. (2009, 2010) and Nadarajah and Zografos (2003).

If we deal with (h, ϕ) -divergence measures in our context, the following can be proved:

Theorem 4 Under the assumptions of Theorem 1, the asymptotic distribution of the family of empirical test statistics defined by

$$S^{\phi_1, h}(\hat{\mathbf{p}}, \mathbf{p}(\hat{\boldsymbol{\lambda}}_{\phi_2}, \hat{\boldsymbol{\eta}}_{\phi_2})) := \frac{2N}{\phi_1''(1)h'(0)} h\left(D_{\phi_1}(\hat{\mathbf{p}}, \mathbf{p}(\hat{\boldsymbol{\lambda}}_{\phi_2}, \hat{\boldsymbol{\eta}}_{\phi_2}))\right)$$

is chi-square with $2^k - (u + t) - 1$ degrees of freedom.

Proof. We have

$$h(x) = h(0) + h'(0)x + o(x) = h'(0)x + o(x),$$

and so

$$h\left(D_{\phi_1}\left(\left(\hat{\mathbf{p}}, \mathbf{p}(\hat{\boldsymbol{\lambda}}_{\phi_2}, \hat{\boldsymbol{\eta}}_{\phi_2})\right)\right)\right) = h'(0)D_{\phi_1}\left(\hat{\mathbf{p}}, \mathbf{p}(\hat{\boldsymbol{\lambda}}_{\phi_2}, \hat{\boldsymbol{\eta}}_{\phi_2})\right) + o\left(D_{\phi_1}\left(\hat{\mathbf{p}}, \mathbf{p}(\hat{\boldsymbol{\lambda}}_{\phi_2}, \hat{\boldsymbol{\eta}}_{\phi_2})\right)\right).$$

Then, the required result follows upon applying Theorem 1. ■

3 Nested latent class models

In the previous example we have seen that the LCM proposed (that we will call M_1) fits our data; however, a question arises: Is it possible to find a latent model with a reduced number of parameters that also fits the data? If the answer is positive, the reduced model should be used instead of M_1 .

Example 5 Consider the example studied in the previous section. In Formann (1985), the following reduced models are studied:

M_2 : Attitudinal changes between the two moments are dependent on the latent classes but are independent on the items.

M_3 : Attitudinal changes between the two moments are independent both on the items and on the latent classes.

M_4 : There are no attitudinal changes.

These different models imply different number of parameters λ_i . More concretely, model M_2 needs six parameters λ_i , model M_3 needs five parameters and finally model M_4 needs four parameters. The corresponding matrices Q_i for these models can be found in Formann (1985).

As for M_1 , $c_{ij} = 0, \forall i, j$ and 4 parameters η_j are considered, taking matrix V as the identity matrix and $d_j = 0, \forall j$.

We can observe that

$$\Theta_{M_1} \supset \Theta_{M_2} \supset \Theta_{M_3} \supset \Theta_{M_4},$$

being Θ_{M_i} the parameter space associated to the LCM M_i . Therefore, we have a nested sequence of LCM.

In general, we shall assume that we have m LCM $\{M_l\}_{l=1, \dots, m}$ in such a way that the parameter space associated to M_l , $l = 1, \dots, m$, is Θ_{M_l} and

$$\Theta_{M_m} \subset \Theta_{M_{m-1}} \subset \dots \subset \Theta_{M_1} \subset \mathbb{R}^t$$

holds. Let us denote $\dim(\Theta_{M_l}) = h_l$; $l = 1, \dots, m$ with

$$h_m < h_{m-1} < \dots < h_1 \leq t,$$

i.e., the parameters of one LCM are a subset of the parameters of the other. Our strategy is to test successively

$$H_{l+1} : \theta \in \Theta_{M_{l+1}} \text{ against } H_l : \theta \in \Theta_{M_l}, \quad l = 1, \dots, m-1. \quad (14)$$

We continue to test as long as the null hypothesis is accepted, and choose the LCM M_l with parameter space Θ_{M_l} according to the first l satisfying that H_{l+1} is rejected (as null hypothesis) in favor of H_l (as alternative hypothesis). This strategy is quite standard for nested models (Cressie et al., 2003). In this section we present two families of phi-divergence test statistics for solving the tests presented in (14).

Let us introduce some additional notation in order to be able to formulate the nested LCM in a convenient way for our purposes. We shall denote by $\theta^A = (\theta^{A,1}, \theta^{A,2}, \theta^{A,3}, \theta^{A,4})$ with $\theta^{A,1} = (\lambda_1, \dots, \lambda_{t^*})$, $\theta^{A,2} = (\lambda_{t^*+1}, \dots, \lambda_t)$, $\theta^{A,3} = (\eta_1, \dots, \eta_{u^*})$ and $\theta^{A,4} = (\eta_{u^*+1}, \dots, \eta_u)$ the parameters associated to the LCM A and by $\theta^B = (\theta^{A,1}, \mathbf{0}, \theta^{A,3}, \mathbf{0})$ the parameters associated to the LCM B . We shall assume that $t + u = h_1$ and $t^* + u^* = h_2$. It is clear that the LCM B is nested in LCM A .

It can be observed that the testing problem given in (14) can be equivalently formulated using the previous notation in the following way:

$$H_{Null} : \theta^{A,2} = \mathbf{0}_{t-t^*} \text{ and } \theta^{A,4} = \mathbf{0}_{u-u^*}. \quad (15)$$

The expression of the classical likelihood ratio test for solving (15) is

$$G_{A-B}^2 = 2 \sum_{\nu=1}^{2^k} n_\nu \log \frac{p(\mathbf{y}_\nu, \hat{\theta}^A)}{p(\mathbf{y}_\nu, \hat{\theta}^B)}. \quad (16)$$

Notice that not only the likelihood ratio test can be used for testing (15); the chi-square test statistic given by

$$X_{A-B}^2 = N \sum_{\nu=1}^{2^k} \frac{\left(p(\mathbf{y}_\nu, \hat{\boldsymbol{\theta}}^A) - p(\mathbf{y}_\nu, \hat{\boldsymbol{\theta}}^B) \right)^2}{p(\mathbf{y}_\nu, \hat{\boldsymbol{\theta}}^B)} \quad (17)$$

can be also used instead.

We can observe that

$$G_{A-B}^2 = 2N \left(D_{Kullback}(\hat{\mathbf{p}}, \mathbf{p}(\hat{\boldsymbol{\theta}}^A)) - D_{Kullback}(\hat{\mathbf{p}}, \mathbf{p}(\hat{\boldsymbol{\theta}}^B)) \right) \quad (18)$$

and

$$X_{A-B}^2 = \frac{2N}{\phi''(1)} D_{Pearson}(\mathbf{p}(\hat{\boldsymbol{\theta}}^A), \mathbf{p}(\hat{\boldsymbol{\theta}}^B)) \quad (19)$$

being $D_{Pearson}(\mathbf{p}(\hat{\boldsymbol{\theta}}^A), \mathbf{p}(\hat{\boldsymbol{\theta}}^B))$ the phi-divergence measure between the probability vectors $\mathbf{p}(\hat{\boldsymbol{\theta}}^A)$ and $\mathbf{p}(\hat{\boldsymbol{\theta}}^B)$ with

$$\phi(x) = \frac{1}{2}(x-1)^2.$$

Based on Eqs. (18) and (19) we are going to give two families of test statistics that are natural extensions of these test statistics for solving the problem of testing given in (14).

A generalization of (18) is obtained if we replace the Kullback-Leibler divergence measure for a phi-divergence measure, i.e.,

$$S_{A-B}^{\phi_1, \phi_2} = \frac{2N}{\phi_1''(1)} \left(D_{\phi_1}(\hat{\mathbf{p}}, \mathbf{p}(\hat{\boldsymbol{\theta}}_{\phi_2}^A)) - D_{\phi_1}(\hat{\mathbf{p}}, \mathbf{p}(\hat{\boldsymbol{\theta}}_{\phi_2}^B)) \right), \quad (20)$$

and a generalization of (19) is achieved if we replace the Pearson divergence measure for a phi-divergence measure, i.e.,

$$T_{A-B}^{\phi_1, \phi_2} = \frac{2N}{\phi_1''(1)} D_{\phi_1}(\mathbf{p}(\hat{\boldsymbol{\theta}}_{\phi_2}^A), \mathbf{p}(\hat{\boldsymbol{\theta}}_{\phi_2}^B)). \quad (21)$$

The previous extensions have been considered in many statistical applications, see for example Cressie et al. (2003), Pardo (2006) and references therein.

In the following theorem we shall obtain the asymptotic distribution of the family of test statistics given in (6) and (6).

Theorem 6 *Given the LCM for binary data A and B with parameters $\boldsymbol{\theta}^A = (\boldsymbol{\theta}^{A,1}, \boldsymbol{\theta}^{A,2}, \boldsymbol{\theta}^{A,3}, \boldsymbol{\theta}^{A,4})$ and $\boldsymbol{\theta}^B = (\boldsymbol{\theta}^{A,1}, \mathbf{0}, \boldsymbol{\theta}^{A,3}, \mathbf{0})$, respectively, and under the null hypothesis given in (15), it follows*

$$S_{A-B}^{\phi_1, \phi_2} \xrightarrow[N \rightarrow \infty]{\mathcal{L}} \chi_{h_1 - h_2}^2,$$

and

$$T_{A-B}^{\phi_1, \phi_2} \xrightarrow[N \rightarrow \infty]{\mathcal{L}} \chi_{h_1 - h_2}^2.$$

Proof. See Appendix. ■

Example 7 (Continuation of Example 1) *We shall consider the sequence of LCM*

$$\boldsymbol{\Theta}_{M_1} \supset \boldsymbol{\Theta}_{M_2} \supset \boldsymbol{\Theta}_{M_3} \supset \boldsymbol{\Theta}_{M_4},$$

In a similar way as in the previous section we consider $\hat{\boldsymbol{\theta}}_{2/3}$ in order to estimate the parameters of the different models. For testing, we consider the family of phi-divergences test statistics $S_{A-B}^{\phi_1, \phi_2}$ and $T_{A-B}^{\phi_1, \phi_2}$ given in (6) and (6), being $\phi_2(x) = \phi_a(x)$, with $a = 2/3$, and $\phi_a(x)$ defined in (12). For $\phi_1(x)$ we shall take $\phi_a(x)$ with $a = -1, -1/2, 0, 2/3, 3/2, 2, 5/2$ and 3. In Table 3 we present the results obtained.

As a conclusion, we can adopt LCM M_2 as the best model in all cases. As before, the values obtained are very similar, due to the asymptotical results.

a/Model	$M_1 - M_2$	$M_2 - M_3$	$M_3 - M_4$		$M_1 - M_2$	$M_2 - M_3$	$M_3 - M_4$
-1	3.761	4.610	31.465		3.431	4.613	31.005
-1/2	3.757	4.593	30.977		3.417	4.604	30.845
0	3.755	4.584	30.769		3.403	4.595	30.722
2/3	3.754	4.578	30.626		3.386	4.585	30.616
1	3.754	4.580	30.659		3.378	4.580	30.587
3/2	3.756	4.586	30.820		3.366	4.574	30.574
2	3.759	4.599	30.991		3.355	4.570	30.597
5/2	3.763	4.617	31.347		3.344	4.566	30.655
3	3.769	4.641	31.765		3.334	4.563	30.749
$\chi^2_{i;0.05}$	5.99	3.84	3.84		5.99	3.84	3.84

Table 3: Results for Example 8 for statistics S (left) and T (right).

Remark 8 Using the ideas given in Remark 3 we can consider the following two families of (h, ϕ) -divergence test statistics:

$$S_{A-B}^{\phi_1, \phi_2, h} = \frac{2N}{\phi_1''(1)h'(0)} \left(h \left(D_{\phi_1} \left(\hat{\mathbf{p}}, \mathbf{p}(\hat{\boldsymbol{\theta}}_{\phi_2}^A) \right) \right) - h \left(D_{\phi_1} \left(\hat{\mathbf{p}}, \mathbf{p}(\hat{\boldsymbol{\theta}}_{\phi_2}^B) \right) \right) \right),$$

and

$$T_{A-B}^{\phi_1, \phi_2, h} = \frac{2N}{\phi_1''(1)h'(0)} h \left(D_{\phi} \left(\mathbf{p}(\hat{\boldsymbol{\theta}}_{\phi_2}^A), \mathbf{p}(\hat{\boldsymbol{\theta}}_{\phi_2}^B) \right) \right).$$

It is easy to establish that again

$$S_{A-B}^{\phi_1, \phi_2, h} \xrightarrow[N \rightarrow \infty]{\mathcal{L}} \chi_{h_1 - h_2}^2,$$

and

$$T_{A-B}^{\phi_1, \phi_2, h} \xrightarrow[N \rightarrow \infty]{\mathcal{L}} \chi_{h_1 - h_2}^2.$$

4 Simulation

Sections 2 and 3 present theoretical results for testing hypothesis in latent models with binary data. These results give the asymptotic distribution theory for the phi-divergence test statistics given in (11), (6) and (6) under the null hypothesis. In this section we present a simulation study to analyze the behavior of this statistics in small samples. We shall analyze the test statistics given in (11).

In Felipe et al. (2014), it was established that the best way to estimate the unknown parameters from the point of view of the efficiency as well as the robustness was the minimum power divergence obtained for $a = 2/3$, as this estimator balances infinitesimal robustness and asymptotic efficiency.

Therefore, in our simulation study we shall consider this estimator. We compare the different test statistics of the family $T_{\Phi_a}^{\Phi_{2/3}}$ defined in (13). The theoretical LCM with binary data that we shall consider in our simulation study is given by a theoretical model with 5 dichotomous questions and 10 latent classes; next, 7 parameters λ_j and 6 parameters η_k are considered; the corresponding matrices of the model are

$$\mathbf{Q}_1 = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix}, \mathbf{Q}_2 = \begin{pmatrix} 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \end{pmatrix}, \mathbf{Q}_3 = \begin{pmatrix} 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 & 0 \end{pmatrix}, \mathbf{Q}_4 = \begin{pmatrix} 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix}$$

$$\mathbf{Q}_5 = \begin{pmatrix} 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \end{pmatrix}, \mathbf{Q}_6 = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \end{pmatrix}, \mathbf{Q}_7 = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix}.$$

Matrix \mathbf{C} is the null matrix. Matrix \mathbf{V} is given by

$$\mathbf{V} = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 & 1 & 0 \end{pmatrix},$$

while $\mathbf{d} = \mathbf{0}$. The theoretical values for vector $\boldsymbol{\lambda}$ and $\boldsymbol{\eta}$ are

$$\boldsymbol{\lambda}_0 = (\lambda_1^0, \dots, \lambda_7^0) = (-3, -2, -1, 0, 1, 2, 3), \boldsymbol{\eta}_0 = (\eta_1^0, \dots, \eta_6^0) = (0.5, 1, 1.5, 2, 2.5, 3).$$

We shall also consider different values of a ; more concretely, we consider $a = -0.5, 0, 2/3, 1$.

For each value of a we consider $R = 10000$ simulations and we reproduce the study for different sample sizes: 200, 300, 400, 500 and 1000. We must not forget that for $a = 0$ and $a = 1$ we have the likelihood ratio test and the chi-square ratio test statistics, respectively, but the unknown parameters are estimated using the minimum power divergence estimator with $a = 2/3$ instead of the maximum likelihood estimator.

We consider as nominal size $\alpha = 0.05$ and compute the *simulated exact size*

$$\hat{\alpha}_n^a := \frac{\#T_{\phi_a}^{\phi_{2/3}} > \chi_{g.l.;0.05}^2}{R}.$$

As explained in Dale (1986), we only consider the test statistics whose simulated exact size $\hat{\alpha}_n^a$ satisfies

$$|\logit(1 - \hat{\alpha}_n^a) - \logit(1 - \alpha)| \leq 0.35 \quad (22)$$

where $\logit(p) = \log(\frac{p}{1-p})$. As a consequence, we only take under consideration the test statistics such that

$$\hat{\alpha}_n^a \in (0.0357, 0.0695). \quad (23)$$

At the same time we obtain the *simulated exact power* for different alternative hypothesis. More concretely, we shall consider a model with a new parameter λ_8 whose corresponding matrix \mathbf{Q}_8 is given by

$$\mathbf{Q}_8 = \begin{pmatrix} 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix},$$

and where this new parameter takes different values, namely -3, -2, -1.5, -1, -0.8, 0, 0.7, 0.9, 1, 1.3, 1.5, 2. Each of these values is related to an alternative hypothesis, except when considering value 0, that corresponds to the null hypothesis.

Simulating observations from each alternative hypothesis we get the simulated exact power for such alternatives

$$\hat{\beta}^a := \frac{\#T_{\phi_a}^{\phi_{2/3}} > \chi_{g.L.;0.05}^2}{R}.$$

In Table 4 we present the simulated exact size as well as the simulated exact power for different values of a .

N	a	-3	-2	-1.5	-1	-0.8	0	0.7	0.9	1	1.3	1.5	2
200	-5	0.7764	0.5242	0.4945	0.4786	0.4510	0.4041	0.5449	0.6352	0.6805	0.8226	0.8937	0.9876
	0	0.4833	0.1662	0.1595	0.1505	0.1455	0.1095	0.2305	0.3232	0.3835	0.6087	0.7498	0.9591
	2/3	0.3181	0.0617	0.0562	0.0472	0.0460	0.0354	0.1047	0.1769	0.2377	0.4668	0.6335	0.9281
	1	0.2915	0.0497	0.0426	0.0338	0.0351	0.0269	0.0885	0.1523	0.2124	0.4363	0.6072	0.9190
300	-5	0.8009	0.4473	0.4090	0.3724	0.3371	0.2680	0.4566	0.5838	0.6604	0.8518	0.9388	0.9987
	0	0.6202	0.1984	0.1789	0.1646	0.1410	0.0946	0.2519	0.3949	0.4891	0.7569	0.8919	0.9958
	2/3	0.5094	0.1041	0.0876	0.0772	0.0627	0.0400	0.1599	0.2911	0.3818	0.6867	0.8534	0.9938
	1	0.4887	0.0870	0.0714	0.0621	0.0515	0.0331	0.1417	0.2687	0.3568	0.6714	0.8438	0.9937
400	-5	0.8311	0.4086	0.3823	0.3195	0.2828	0.1933	0.4200	0.5925	0.6811	0.9107	0.9723	0.9999
	0	0.7337	0.2245	0.2028	0.1712	0.1493	0.0819	0.2871	0.4761	0.5878	0.8739	0.9603	0.9997
	2/3	0.6670	0.1429	0.1222	0.1006	0.0863	0.0429	0.2124	0.4027	0.5214	0.8444	0.9501	0.9996
	1	0.6531	0.1274	0.1099	0.0870	0.0733	0.0360	0.1995	0.3855	0.5052	0.8368	0.9487	0.9996
500	-5	0.8793	0.4129	0.3626	0.3065	0.2578	0.1448	0.4212	0.6309	0.7361	0.9461	0.9893	1.0000
	0	0.8207	0.2625	0.2293	0.2015	0.1626	0.0743	0.3357	0.5598	0.6781	0.9320	0.9864	1.0000
	2/3	0.7821	0.1909	0.1600	0.1344	0.1075	0.0448	0.2755	0.5064	0.6388	0.9204	0.9836	1.0000
	1	0.7748	0.1769	0.1458	0.1196	0.0943	0.0397	0.2623	0.4962	0.6289	0.9177	0.9833	1.0000
1000	-5	0.9924	0.4910	0.5226	0.4639	0.3192	0.0817	0.6081	0.8825	0.9537	0.9999	1.0000	1.0000
	0	0.9916	0.3216	0.3498	0.3368	0.2781	0.0654	0.5854	0.8750	0.9506	0.9999	1.0000	1.0000
	2/3	0.9910	0.2356	0.2633	0.2528	0.2399	0.0510	0.5681	0.8723	0.9487	0.9998	1.0000	1.0000
	1	0.9913	0.2171	0.2431	0.2303	0.2285	0.0479	0.5650	0.8717	0.9487	0.9999	1.0000	1.0000

Table 4: Exact level and power for different values of N and a .

We also present the pictures for each sample size of the different alternative hypothesis for the test statistic $\lambda = -1/2, 0, 2/3, 1$ in Figures 1 to 5.

As it can be observed in Table 4 (see the column corresponding to 0) and Figures 1 to 5, the simulated level is outside the interval given in (23) for $a = 0, -0.5$ for all sample sizes under consideration; besides, for sample sizes $N = 400, 500, 1000$, the test statistic corresponding to $a = 1$ lays inside this interval. Notice that the test statistic for $a = 2/3$ is the only one laying in this interval for any sample size. As a straightforward conclusion, the test statistic for $a = 2/3$ seems to be the best one for sample sizes $N = 200, 300$, and we just need to choose between $a = 1$ and $a = 2/3$ for $N = 400, 500, 1000$. For making this decision, we focus on the simulated power values, noting that they are higher for $a = 2/3$ than for $a = 1$; we then conclude that $a = 2/3$ seems to show



Figure 1: Simulated exact level and power for $N = 200$.

a better behavior than the likelihood ratio test statistic and Pearson test statistic (with estimations obtained through $a = 2/3$ instead of maximum likelihood) when dealing with LCM for binary data.

5 Conclusions

In this paper we have introduced phi-divergence test statistics in the context of LCM for binary data. In a previous paper, we have already shown that phi-divergence estimators can be a useful tool in this framework; now, we have treated two new problems: the problem of goodness-of-fit and the problem of selecting the best model throughout a nested sequence of models. Classically, as it can be seen for instance in Formann (1985), these problems have been considered on the basis of the likelihood-ratio-test and the chi-square test statistic. In both of them, we have derived two families of test statistics generalizing the classical ones; besides, we have obtained their asymptotical distribution under the null hypothesis of that LCM fits the data, showing that it coincides with the one of the classical test statistics; thus, they show the same behavior as the classical statistics for big sample sizes.

At this point, an interesting problem arises: are there differences for small or moderate sample sizes? To deal with this problem, we have carried out a simulation study; from this study, it seems that the phi-divergence test statistic for $a = 2/3$ shows a better behavior than the classical test statistics.

6 Acknowledgements

This work was partially supported by Grant MTM2012-33740.

7 Appendix

Proof of Theorem 1

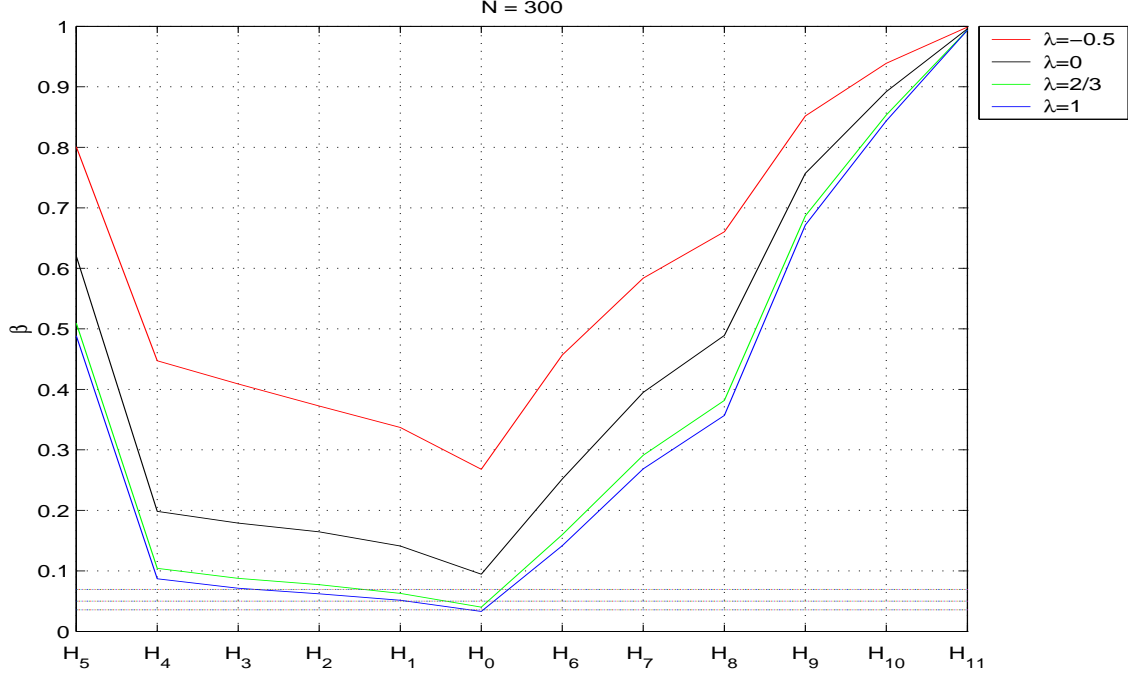


Figure 2: Simulated exact level and power for $N = 300$.

A second-order Taylor expansion of $D_{\phi_1}(\mathbf{p}, \mathbf{q})$ around $(\mathbf{p}(\theta_0), \mathbf{p}(\theta_0))$ at $(\hat{\mathbf{p}}, \mathbf{p}(\hat{\theta}_{\phi_2}))$ is given by

$$D_{\phi_1}(\hat{\mathbf{p}}, \mathbf{p}(\hat{\theta}_{\phi_2})) = \frac{\phi_1''(1)}{2} (\hat{\mathbf{p}} - \mathbf{p}(\hat{\theta}_{\phi_2}))^T \mathbf{D}_{\mathbf{p}(\theta_0)}^{-1} (\hat{\mathbf{p}} - \mathbf{p}(\hat{\theta}_{\phi_2})) + o_p(N^{-1}),$$

being $\theta_0 = (\lambda_1^0, \dots, \lambda_u^0, \eta_1^0, \dots, \eta_s^0)$. By $\mathbf{D}_{\mathbf{p}(\theta_0)}$ we are denoting the diagonal matrix with $\mathbf{p}(\theta_0)$ in the main diagonal. By Theorem 1 in Felipe et al (2014) we have

$$\hat{\theta}_{\phi_2} - \theta_0 = \left(\mathbf{L}(\theta_0)^T \mathbf{L}(\theta_0) \right)^{-1} \mathbf{L}(\theta_0)^T \mathbf{D}_{\mathbf{p}(\theta_0)}^{-1/2} (\hat{\mathbf{p}} - \mathbf{p}(\theta_0)) + o_p(N^{-1/2}),$$

with

$$\mathbf{L}(\theta_0) = \mathbf{D}_{\mathbf{p}(\theta_0)}^{-1/2} \left(\frac{\partial \mathbf{p}(\theta)}{\partial \theta} \right)_{\theta=\theta_0}.$$

Therefore,

$$\begin{aligned} \mathbf{p}(\hat{\theta}_{\phi_2}) - \mathbf{p}(\theta_0) &= \left(\frac{\partial \mathbf{p}(\theta)}{\partial \theta} \right)_{\theta=\theta_0} (\hat{\theta}_{\phi_2} - \theta_0) + o_p(N^{-1/2}) \\ &= \mathbf{D}_{\mathbf{p}(\theta_0)}^{1/2} \mathbf{L}(\theta_0) \left(\mathbf{L}(\theta_0)^T \mathbf{L}(\theta_0) \right)^{-1} \mathbf{L}(\theta_0)^T \mathbf{D}_{\mathbf{p}(\theta_0)}^{-1/2} (\hat{\mathbf{p}} - \mathbf{p}(\theta_0)) + o_p(N^{-1/2}) \\ &= \mathbf{V}(\theta_0) (\hat{\mathbf{p}} - \mathbf{p}(\theta_0)) + o_p(N^{-1/2}) \end{aligned}$$

with $\mathbf{V}(\theta_0) := \mathbf{D}_{\mathbf{p}(\theta_0)}^{1/2} \mathbf{L}(\theta_0) \left(\mathbf{L}(\theta_0)^T \mathbf{L}(\theta_0) \right)^{-1} \mathbf{L}(\theta_0)^T \mathbf{D}_{\mathbf{p}(\theta_0)}^{-1/2}$.

On the other hand,

$$\sqrt{N} (\hat{\mathbf{p}} - \mathbf{p}(\theta_0)) \xrightarrow[N \rightarrow \infty]{\mathcal{L}} \mathcal{N}(\mathbf{0}, \Sigma_{\mathbf{p}(\theta_0)})$$

being

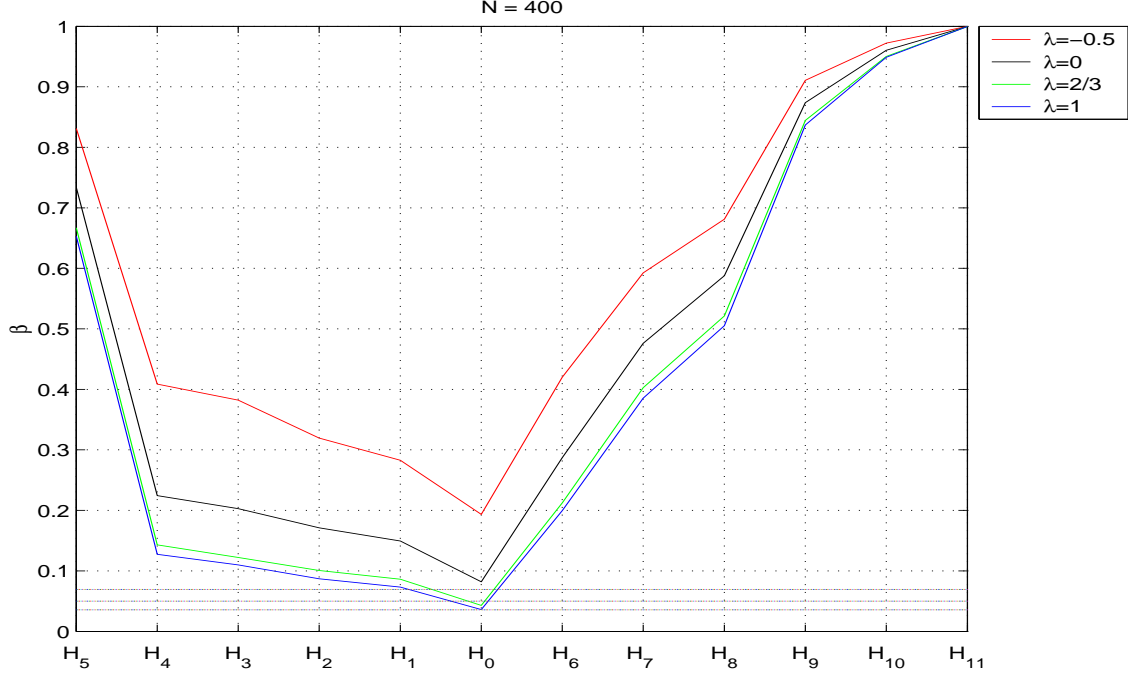


Figure 3: Simulated exact level and power for $N = 400$.

$$\Sigma_{\mathbf{p}(\theta_0)} = \mathbf{D}_{\mathbf{p}(\theta_0)} - \mathbf{p}(\theta_0)\mathbf{p}(\theta_0)^T.$$

Then we have

$$\hat{\mathbf{p}} - \mathbf{p}(\hat{\theta}_{\phi_2}) = (\mathbf{I} - \mathbf{V}(\theta_0))(\hat{\mathbf{p}} - \mathbf{p}(\theta_0)) + o_p(N^{-1/2}),$$

and we conclude that

$$\sqrt{N}(\hat{\mathbf{p}} - \mathbf{p}(\hat{\theta}_{\phi_2})) \xrightarrow[N \rightarrow \infty]{\mathcal{L}} \mathcal{N}\left(\mathbf{0}, (\mathbf{I} - \mathbf{V}(\theta_0)^T) \Sigma_{\mathbf{p}(\theta_0)} (\mathbf{I} - \mathbf{V}(\theta_0)^T)\right).$$

Notice that the asymptotic distribution of

$$\frac{2N}{\phi_1''(1)} D_{\phi_1}(\hat{\mathbf{p}}, \mathbf{p}(\hat{\theta}_{\phi_2}))$$

coincides with the asymptotic distribution of the quadratic form

$$N(\hat{\mathbf{p}} - \mathbf{p}(\hat{\theta}_{\phi_2}))^T \mathbf{D}_{\mathbf{p}(\theta_0)}^{-1} (\hat{\mathbf{p}} - \mathbf{p}(\hat{\theta}_{\phi_2})) = \sqrt{N}(\hat{\mathbf{p}} - \mathbf{p}(\hat{\theta}_{\phi_2}))^T \mathbf{D}_{\mathbf{p}(\theta_0)}^{-1/2} \mathbf{D}_{\mathbf{p}(\theta_0)}^{-1/2} (\hat{\mathbf{p}} - \mathbf{p}(\hat{\theta}_{\phi_2})) \sqrt{N} = X^T X,$$

with

$$X := \sqrt{N} \mathbf{D}_{\mathbf{p}(\theta_0)}^{-1/2} (\hat{\mathbf{p}} - \mathbf{p}(\hat{\theta}_{\phi_2})).$$

Now, as

$$X \xrightarrow[N \rightarrow \infty]{\mathcal{L}} \mathcal{N}\left(\mathbf{0}, \mathbf{D}_{\mathbf{p}(\theta_0)}^{-1/2} (\mathbf{I} - \mathbf{V}(\theta_0)^T) \Sigma_{\mathbf{p}(\theta_0)} (\mathbf{I} - \mathbf{V}(\theta_0)^T) \mathbf{D}_{\mathbf{p}(\theta_0)}^{-1/2}\right),$$

we conclude that the asymptotic distribution of $X^T X$ will be a chi-square distribution if the matrix

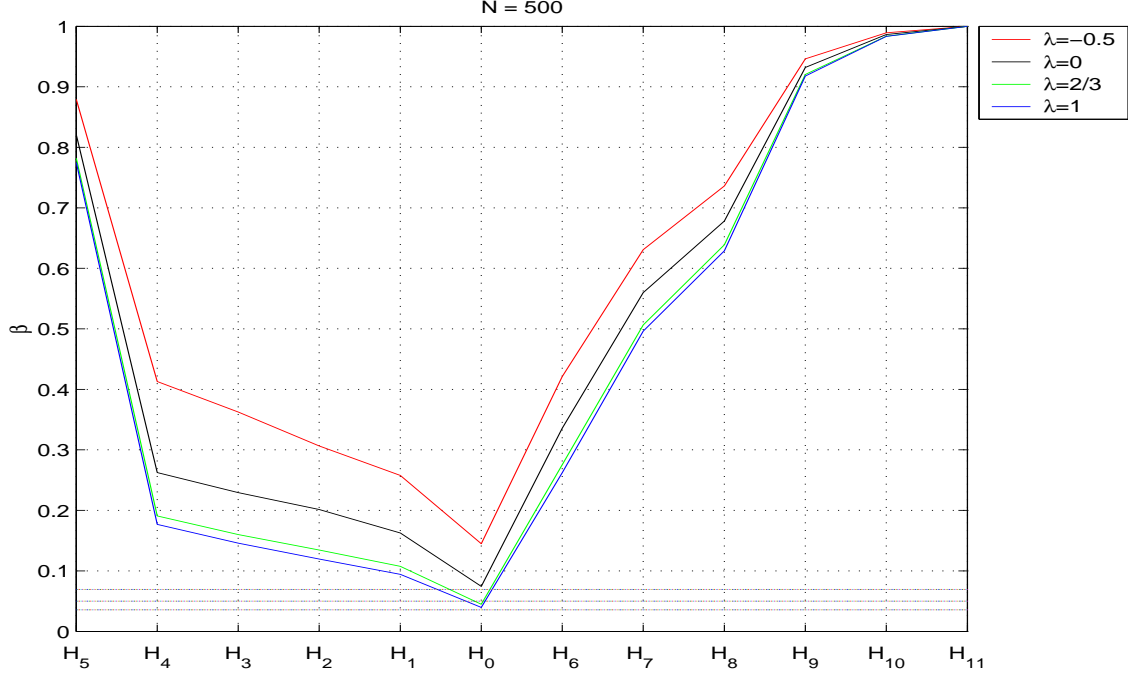


Figure 4: Simulated exact level and power for $N = 500$.

$$\mathbf{Q}(\theta_0) := D_{\mathbf{p}(\theta_0)}^{-1/2} \left(I - V(\theta_0)^T \right) \Sigma_{\mathbf{p}(\theta_0)} \left(I - V(\theta_0)^T \right) D_{\mathbf{p}(\theta_0)}^{-1/2}$$

is idempotent and symmetric, and in this case the degrees of freedom will be the trace of the matrix $\mathbf{Q}(\theta_0)$. Symmetry is evident. Establishing that the matrix $\mathbf{Q}(\theta_0)$ is idempotent and that its trace is $2^k - (u + t) - 1$ is a simple but long and tedious exercise; a detailed proof of this fact can be found in Pardo (2006) (Theorem 6.1, pag. 259).

Proof of Theorem 6

Based on Theorem 1 in Felipe et al. (2014) we have

$$\hat{\theta}_{\phi_2}^A - \theta_0^A = \left(L(\theta_0^A)^T L(\theta_0^A) \right)^{-1} L(\theta_0^A)^T D_{\mathbf{p}(\theta_0^A)}^{-1/2} (\hat{\mathbf{p}} - \mathbf{p}(\theta_0^A)) + o_p(N^{-1/2}),$$

being

$$\theta_0^A = (\theta_0^{A,1}, 0, \theta_0^{A,3}, 0), \quad \mathbf{p}(\theta_0^A) = (p(y_1, \theta_0^A), \dots, p(y_{2^k}, \theta_0^A))$$

and

$$L(\theta_0^A) = D_{\mathbf{p}(\theta_0^A)}^{-1/2} \left(\frac{\partial \mathbf{p}(\theta^A)}{\partial \theta^{A,1}}, \frac{\partial \mathbf{p}(\theta^A)}{\partial \theta^{A,2}}, \frac{\partial \mathbf{p}(\theta^A)}{\partial \theta^{A,3}}, \frac{\partial \mathbf{p}(\theta^A)}{\partial \theta^{A,4}} \right)_{\theta^A = \theta_0^A}.$$

Similarly,

$$\hat{\theta}_{\phi_2}^B - \theta_0^B = \left(M(\theta_0^B)^T M(\theta_0^B) \right)^{-1} M(\theta_0^B)^T D_{\mathbf{p}(\theta_0^B)}^{-1/2} (\hat{\mathbf{p}} - \mathbf{p}(\theta_0^B)) + o_p(N^{-1/2})$$

being

$$M(\theta_0^B) = D_{\mathbf{p}(\theta_0^B)}^{-1/2} \left(\frac{\partial \mathbf{p}(\theta^B)}{\partial \theta^{B,1}}, \frac{\partial \mathbf{p}(\theta^B)}{\partial \theta^{B,3}} \right)_{\theta^B = \theta_0^B}.$$

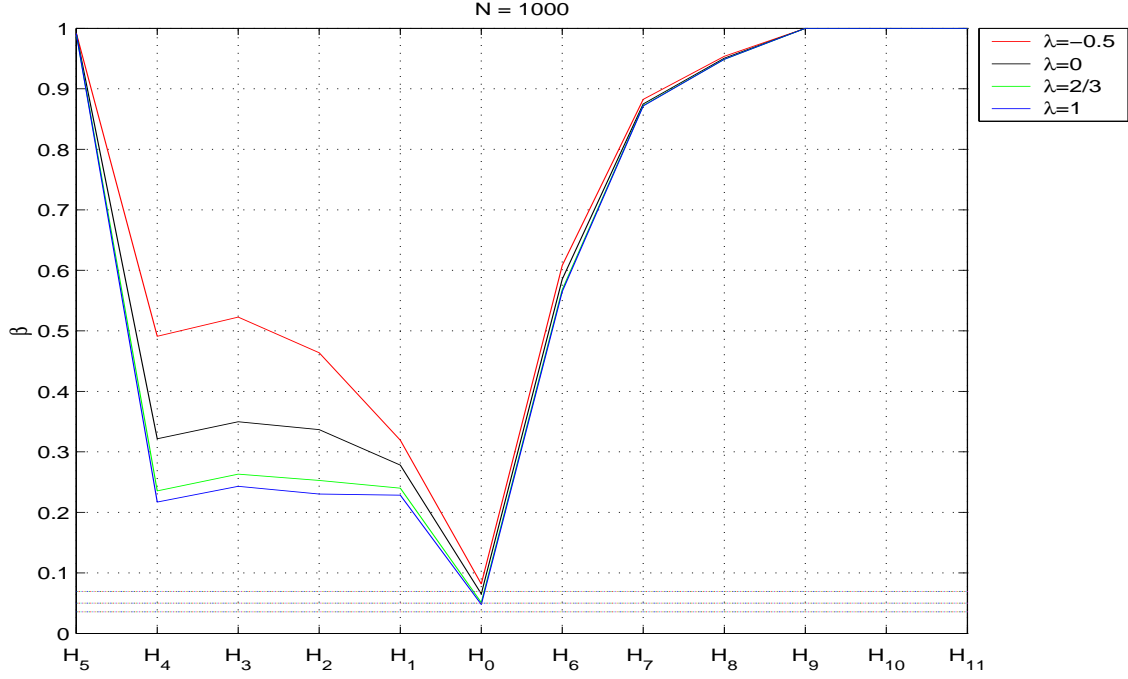


Figure 5: Simulated exact level and power for $N = 1000$.

As

$$\theta_0^B = (\theta_0^{A,1}, \theta_0^{A,3}),$$

and by the hypothesis,

$$(\theta_0^{A,2}, \theta_0^{A,4}) = \mathbf{0},$$

it follows that

$$\hat{\theta}_{\phi_2}^B = (\hat{\theta}_{\phi_2}^{A,1}, \hat{\theta}_{\phi_2}^{A,3}), \theta_0^B = (\theta_0^{A,1}, \theta_0^{A,3})$$

and

$$p(\theta_0^B) = p(\theta_0^A),$$

whence

$$D_{p(\theta_0^B)} = D_{p(\theta_0^A)}, \quad M(\theta_0^B) = M(\theta_0^A) = D_{p(\theta_0^A)}^{-1/2} \left(\frac{\partial p(\theta^A)}{\partial \theta^{A,1}}, \frac{\partial p(\theta^A)}{\partial \theta^{A,3}} \right)_{\theta^A = \theta_0^A}.$$

Therefore,

$$p(\hat{\theta}_{\phi_2}^A) - p(\theta_0^A) = D_{p(\theta_0^A)}^{1/2} L(\theta_0^A) \left(L(\theta_0^A)^T L(\theta_0^A) \right)^{-1} L(\theta_0^A)^T D_{p(\theta_0^A)}^{-1/2} (\hat{p} - p(\theta_0^A)) + o_p(N^{-1/2})$$

and

$$p(\hat{\theta}_{\phi_2}^B) - p(\theta_0^B) = p(\hat{\theta}_{\phi_2}^B) - p(\theta_0^A) = D_{p(\theta_0^A)}^{1/2} M(\theta_0^A) \left(M(\theta_0^A)^T M(\theta_0^A) \right)^{-1} M(\theta_0^A)^T D_{p(\theta_0^A)}^{-1/2} (\hat{p} - p(\theta_0^A)) + o_p(N^{-1/2}).$$

Then,

$$D_{p(\theta_0^A)}^{-1/2} (p(\hat{\theta}_{\phi_2}^A) - p(\hat{\theta}_{\phi_2}^B)) = (R_L(\theta_0^A) - R_M(\theta_0^A)) D_{p(\theta_0^A)}^{-1/2} (\hat{p} - p(\theta_0^A)) + o_p(N^{-1/2})$$

with

$$\mathbf{R}_L(\boldsymbol{\theta}_0^A) = \mathbf{L}(\boldsymbol{\theta}_0^A) \left(\mathbf{L}(\boldsymbol{\theta}_0^A)^T \mathbf{L}(\boldsymbol{\theta}_0^A) \right)^{-1} \mathbf{L}(\boldsymbol{\theta}_0^A)^T, \quad \mathbf{R}_M(\boldsymbol{\theta}_0^A) = \mathbf{M}(\boldsymbol{\theta}_0^A) \left(\mathbf{M}(\boldsymbol{\theta}_0^A)^T \mathbf{M}(\boldsymbol{\theta}_0^A) \right)^{-1} \mathbf{M}(\boldsymbol{\theta}_0^A)^T.$$

Therefore the asymptotic distribution of

$$\sqrt{N} \mathbf{D}_{\mathbf{p}(\boldsymbol{\theta}_0^A)}^{-1/2} \left(\mathbf{p}(\hat{\boldsymbol{\theta}}_{\phi_2}^A) - \mathbf{p}(\tilde{\boldsymbol{\theta}}_{\phi_2}) \right)$$

is a normal distribution with vector mean zero and variance-covariance matrix

$$\boldsymbol{\Sigma}_{AB} = \left(\mathbf{R}_L(\boldsymbol{\theta}_0^A) - \mathbf{R}_M(\boldsymbol{\theta}_0^A) \right) \mathbf{D}_{\mathbf{p}(\boldsymbol{\theta}_0^A)}^{-1/2} \boldsymbol{\Sigma}_{\mathbf{p}(\boldsymbol{\theta}_0^A)} \mathbf{D}_{\mathbf{p}(\boldsymbol{\theta}_0^A)}^{-1/2} \left(\mathbf{R}_L(\boldsymbol{\theta}_0^A) - \mathbf{R}_M(\boldsymbol{\theta}_0^A) \right),$$

being

$$\boldsymbol{\Sigma}_{\mathbf{p}(\boldsymbol{\theta}_0^A)} = \text{diag}(\mathbf{p}(\boldsymbol{\theta}_0^A)) - \mathbf{p}(\boldsymbol{\theta}_0^A) \mathbf{p}(\boldsymbol{\theta}_0^A)^T$$

It can be established that $\boldsymbol{\Sigma}_{AB}$ can be written as $\boldsymbol{\Sigma}_{AB} = \mathbf{R}_L(\boldsymbol{\theta}_0^A) - \mathbf{R}_M(\boldsymbol{\theta}_0^A)$ because $\mathbf{R}_L(\boldsymbol{\theta}_0^A)$ and $\mathbf{R}_M(\boldsymbol{\theta}_0^A)$ are orthogonal projections operators and the columns of $\mathbf{M}(\boldsymbol{\theta}_0^A)$ are a subset of the columns of $\mathbf{L}(\boldsymbol{\theta}_0^A)$ (see again Pardo (2006), Th. 7.1, pag. 311 for details). Then

$$\mathbf{R}_L(\boldsymbol{\theta}_0^A) \mathbf{R}_M(\boldsymbol{\theta}_0^A) = \mathbf{R}_M(\boldsymbol{\theta}_0^A) \mathbf{R}_L(\boldsymbol{\theta}_0^A) = \mathbf{R}_M(\boldsymbol{\theta}_0^A).$$

At the same time

$$\mathbf{p}(\boldsymbol{\theta}_0^A)^{1/2} \mathbf{R}_L(\boldsymbol{\theta}_0^A) = \mathbf{p}(\boldsymbol{\theta}_0^A)^{1/2} \mathbf{R}_M(\boldsymbol{\theta}_0^A) = \mathbf{0}.$$

Then we have that the matrix $\mathbf{R}_L(\boldsymbol{\theta}_0^A) - \mathbf{R}_M(\boldsymbol{\theta}_0^A)$ is symmetric and idempotent and in this case the number of eigenvalues different to zero and equal 1 coincide with the trace of $\mathbf{R}_L(\boldsymbol{\theta}_0^A) - \mathbf{R}_M(\boldsymbol{\theta}_0^A)$,

$$\begin{aligned} \text{trace}(\mathbf{R}_L(\boldsymbol{\theta}_0^A)) &= \text{trace} \left(\mathbf{L}(\boldsymbol{\theta}_0^A) \left(\mathbf{L}(\boldsymbol{\theta}_0^A)^T \mathbf{L}(\boldsymbol{\theta}_0^A) \right)^{-1} \mathbf{L}(\boldsymbol{\theta}_0^A)^T \right) \\ &= \text{trace} \left(\mathbf{L}(\boldsymbol{\theta}_0^A)^T \mathbf{L}(\boldsymbol{\theta}_0^A) \left(\mathbf{L}(\boldsymbol{\theta}_0^A)^T \mathbf{L}(\boldsymbol{\theta}_0^A) \right)^{-1} \right) \\ &= \text{trace}(\mathbf{I}_{h_1 \times h_2}) = h_1. \end{aligned}$$

Similarly,

$$\text{trace}(\mathbf{R}_M(\boldsymbol{\theta}_0^A)) = h_2.$$

Finally, the second-order expansion of

$$D_{\phi_1}(\mathbf{p}(\hat{\boldsymbol{\theta}}_{\phi_2}^A), \mathbf{p}(\hat{\boldsymbol{\theta}}_{\phi_2}^B))$$

about $(\mathbf{p}(\boldsymbol{\theta}_0^A), \mathbf{p}(\boldsymbol{\theta}_0^A))$ gives

$$\begin{aligned} D_{\phi_1}(\mathbf{p}(\hat{\boldsymbol{\theta}}_{\phi_2}^A), \mathbf{p}(\hat{\boldsymbol{\theta}}_{\phi_2}^B)) &= \frac{\phi_1''(1)}{2} \left(\mathbf{p}(\hat{\boldsymbol{\theta}}_{\phi_2}^A) - \mathbf{p}(\hat{\boldsymbol{\theta}}_{\phi_2}^B) \right)^T \mathbf{D}_{\mathbf{p}(\boldsymbol{\theta}_0^A)}^{-1} \left(\mathbf{p}(\hat{\boldsymbol{\theta}}_{\phi_2}^A) - \mathbf{p}(\hat{\boldsymbol{\theta}}_{\phi_2}^B) \right) + o_p(1) \\ &= \frac{\phi_1''(1)}{2} \left(\mathbf{p}(\hat{\boldsymbol{\theta}}_{\phi_2}^A) - \mathbf{p}(\hat{\boldsymbol{\theta}}_{\phi_2}^B) \right)^T \mathbf{D}_{\mathbf{p}(\boldsymbol{\theta}_0^A)}^{-1/2} \mathbf{D}_{\mathbf{p}(\boldsymbol{\theta}_0^A)}^{-1/2} \left(\mathbf{p}(\hat{\boldsymbol{\theta}}_{\phi_2}^A) - \mathbf{p}(\hat{\boldsymbol{\theta}}_{\phi_2}^B) \right) + o_p(1) \end{aligned}$$

Therefore, as we have shown in the previous proof, the asymptotic distribution of $T_{A-B}^{\phi_1, \phi_2}$ is a chi-square distribution with $h_1 - h_2$ degrees of freedom.

References

- [1] Bhattacharyya, A. (1943). On a measure of divergence between two statistical populations defined by their probability distributions. *Bulletin of the Calcutta Mathematical Society*, 35, 99–109.
- [2] Clogg, C. (1995). Latent class models: Recent developments and prospects for the future. In Arminger, C. G. and Sobol, M., editors, *Handbook of statistical modeling for the social and behavioral sciences*, pages 311–352. Plenum, New York (USA).
- [3] Coleman, J. S. (1964). *Introduction to Mathematical Sociology*. Free Press, New York (USA).
- [4] Cressie, N. and Pardo, L. (2002). Phi-divergence statistics. In: Elshaarawi, A.H., Plegorich, W.W. editors. *Encyclopedia of environmetrics*, vol. 13. pp: 1551–1555, John Wiley and sons, New York.
- [5] Cressie, N., Pardo, L. and Pardo, M. C. (2003). Size and power considerations for testing loglinear models using ϕ -divergence test statistics. *Statistica Sinica*, 13 (2), 555–570.
- [6] Cressie, N. and Read, T. R. C. (1984). Multinomial goodness-of-fit tests. *J. Roy. Statist. Soc. Ser. B*, 8:440–464.
- [7] Felipe, A., Miranda, P. and L. Pardo (2014). Minimum ϕ -divergence estimation in constrained latent class models for binary data. <http://arxiv.org/abs/1406.0109>
- [8] Formann, A. (1976). Schätzung der Parameter in Lazarsfeld Latent-Class Analysis. In *Res. Bull.*, number 18. Institut für Psychologie der Universität Wien. In German.
- [9] Formann, A. (1977). Log-linear Latent Class Analyse. In *Res. Bull.*, number 20. Institut für Psychologie der Universität Wien. In German.
- [10] Formann, A. (1978). A note on parametric estimation for Lazarsfeld’s latent class analysis. *Psychometrika*, 48:123–126.
- [11] Formann, A. (1982). Linear logistic latent class analysis. *Biometrical Journal*, 24:171–190.
- [12] Formann, A. (1985). Constrained latent class models: Theory and applications. *British Journal of Mathematics and Statistical Psychology*, 38:87–111.
- [13] Gil, M., Alajaji, F. and Linder, T. (2013). Rényi divergence measures for commonly used univariate continuous distributions. *Information Sciences*, 249, 124–131.
- [14] Golshani, L. and Pasha, E. (2010). Rényi entropy rate for Gaussian processes. *Information Sciences*, 180, 8, 1486–1491.
- [15] Golshani, L. , Pasha, E. and Yari, G. (2009). Some properties of Rényi entropy and Rényi entropy rate. *Information Sciences*, 179, 14, 2426–2433.
- [16] Goodman, L. A. (1974). Exploratory latent structure analysis using both identifiable and unidentifiable models. *Biometrika*, 61:215–231.
- [17] Lazarsfeld, P. and Henry, N. (1968). *Latent structure analysis*. Houghton-Mifflin, Boston (USA).
- [18] McHugh, R. (1956). Efficient estimation and local identification in Latent Class Analysis. *Psychometrika*, 21:331–347.
- [19] Menéndez, M. L., Morales, D., Pardo, L. and Salicrú, M. (1995). Asymptotic behavior and statistical applications of divergence measures in multinomial populations: A unified study. *Statistical Papers*, **36**, 1–29.
- [20] Menéndez, M. L., Pardo, J. A., Pardo, L. and Pardo, M. C. (1997). Asymptotic approximations for the distributions of the (h, ϕ) -divergence goodness-of-fit statistics: Applications to Rényi’s statistic. *Kybernetes*, **26**, 442–452.

- [21] Morales, D., Pardo, L., and Vajda, I. (1995). Asymptotic divergence of estimators of discrete distributions. *Journal of Statistical Planning and Inference*, 48:347–369.
- [22] Nadarajah, S. and Zografos, K. (2003). Formulas for Rényi information and related measures for univariate distributions. *Information Sciences*, 155, 1-2, 119-138.
- [23] Pardo, L. (2006). *Statistical inference based on divergence measures*. Chapman & Hall/CRC, Boca Raton.
- [24] Rényi, A. (1961). On measures of entropy and information. *Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability*, **1**, 547-5
- [25] Sharma, B. D. and Mittal, D. P. (1997). New non-additive measures of relative information. *Journal of Combinatorics, Information & Systems Science*, **2**, 122-133.